Preprint version prior to publication

Theory and Learning Analytics

1. Theory and learning analytics, a historical perspective

Jason M. Lodge

School of Education, The University of Queensland
St. Lucia, Queensland Australia

ORCID: [0000-0001-6330-6160]

Simon Knight

TD School, University of Technology Sydney
Ultimo, New South Wales, Australia

ORCID: [0000-0002-8709-5780]

Kirsty Kitto

Connected Intelligence Centre, University of Technology Sydney
Ultimo, New South Wales, Australia

ORCID: 0000-0001-7642-7121

Abstract

The first decade of research and thinking in learning analytics has seen shifting foci and evolving theoretical foundations. Indeed, the very role of theory in, about, and of learning analytics has been addressed in different ways across sub-sections of the field. From an early emphasis on data, computing and systems, the field has increasingly connected with theories and ideas from educational research, sociology, philosophy, and the learning sciences. The richness resulting from this confluence of theories provides a foundation for enhancing the use of data and analytics for learning, differentiating learning analytics from other pre-existing fields, and for deepening the understanding of how learning works. However, despite the broadening scope of theoretical perspectives in, about, and of learning analytics, old tensions remain, and new ones have emerged. As is evident in other areas of educational research, there are intractable differences in fundamental philosophies that create barriers to meaningful dialogue and the progression of the field. In this chapter, we will provide an overview of the key theoretical trends in learning analytics research and place these trends within a broader perspective. Specifically, we will describe the theorising of learning analytics, theory in learning analytics, and theories about learning analytics.

- Theoretical origins of learning analytics in historical context
- Theoretical critique from within learning analytics
- Theoretical critique from without learning analytics

Released as a self-archived pre-print version as per Springer Nature Accepted Manuscripts Terms of Use. Readers should refer to the full published version: Lodge, J. M., Knight, S. & Kitto, K. (2024). Theory and learning analytics, a historical perspective. In Bartimote, K., Gasevic, D. & Howard, S. (eds.), *Theory informing and arising from learning analytics.* Springer.

- Theory used in learning analytics
- Theory arising from learning analytics

1.1 Theoretical origins of learning analytics as a field in historical context

The field of learning analytics emerged from a confluence of existing activity and ideas (Elias, 2011), coalescing at the first Learning Analytics and Knowledge (LAK) conference in 2011 (Siemens, 2013). This emergence of the field of learning analytics was spurred by the potential of big data and associated analytic techniques particularly in achieving learning impact at scale in open learning (Siemens, 2013). The field continued a tradition that dates back several decades incorporating evolving technologies and techniques for computing, analysis, and modelling of data. As Rosé (2018) points out, learning analytics emerged during a period of 'strong empiricism' where there was a prevailing view that questions could be effectively answered simply through the collection and analysis of larger pools of data (e.g. 'The End of Theory', Anderson, 2008). An explosion of available data from and about learners, with associated increases in power and effect size was facilitated through the increased use of various technologies in education. Learning Management Systems (LMSs) and other Educational Technology (EdTech) tools created the fertile conditions for the emergence of learning analytics. Understandably there was great excitement about what was possible with this new data and the methods and analytical processes that were emerging from computer and data science, statistics, and mathematical modelling. In parallel, theorising of the emerging learning analytics field was engaged in attempts to position it with respect to other data-informed fields of education (Siemens & Baker, 2012; Balacheff & Lund, 2013), and the growing interest in analytics across business contexts (Ferguson, 2012).

Given this foundation, it is perhaps not surprising that the initial emphasis of the field could fairly be described as focusing on emerging technologies and data availability, and opportunities provided by these (Lang et al., 2022). Indeed, we might make use of an analogy of a clock for much of the work completed during this period: there was a lot of emphasis on telling the time (i.e. reading the data that could be found) and much less questioning about how the clock worked or what perceptions of time even mean. This led to early questioning of the theoretical underpinning of learning analytics, asking if we were measuring what we valued, or simply valuing that which was easy to measure (Knight & Buckingham Shum 2017); or what could be done over what should be done. It would be an overstatement to claim that the earliest forays into learning analytics were atheoretical. However, there have been many who have claimed that the field lacked the theoretical foundation needed for real impact to be made (e.g. Gašević et al., 2016; Wise & Shaffer, 2015).

In the early years following the first LAK conference, several key issues emerged that necessitated a renewed emphasis on the core approaches and missions of learning analytics. Among these issues was what learning analytics is and how it differs from existing fields. Much of the early work in learning analytics took on an applied or pragmatic focus, however, the application of data science to education does not occur in a vacuum and the conceptualisation, measurement and inferences that are made require substantial theorisation (Gašević et al., 2016). This became apparent as the research on the use of data to make predictions about student achievement began to flourish, and different models were found to make very different assumptions about cohorts, their socio-economic status, and even what was meant by success. Thus, this early *theorising of* learning analytics as a distinct field, community, or domain of application (i.e., data science, applied to education), sat alongside early empiricism and pragmatism. Over the history of the field, this theorising has continued, with discussion about how learning analytics is aligned and different from other areas such as computer-supported collaborative learning (CSCL), artificial intelligence in education (AIED), and perhaps most prominently, educational data mining (EDM). Theories in this grouping are focussed on what learning analytics is. As the field has evolved, while aspects of these discussions have settled, we nevertheless see continued evolution in the definition of 'learning analytics'.

None of this is to say that there does not remain substantial theorising about learning analytics to be done. While many are satisfied with the definition adopted by The Society for Learning Analytics Research (SoLAR, n.d.), there remains some disagreement about what the field is (e.g. Macfadyen et al., 2020), and this is reflected in the number of alternative field definitions in circulation. For example, when describing the difference between learning analytics and the cognate field of educational data mining, Rienties et al. (2020) argue that the defining characteristic of learning analytics is the aim of improving learning processes. Questions about what the field is are important given the ethical and moral concerns of many in the learning analytics community. If, for example, the purpose is to uncover new insights into learning, this definition suggests a rather different focus than to provide insights regarding learning to students and teachers, or addressing historic injustices and inequity, or the tuning of machine learning models. There are tensions in scale (number of learners versus effect size), evaluation (accuracy versus effect size), and audience, with a range of implications for data rights, transparency, anonymity, and ethics (Kitto & Knight, 2019). Theory of learning analytics as a sharply defined field and discipline, has thus morphed into broader discussions regarding the purpose of the field.

1.2 Theoretical critique from *within* learning analytics

As we have described, much early work in learning analytics focussed on attempting to make predictions about student progress using data. The early days of learning analytics saw much work focussing on improving the retention of students in university contexts (Colvin et al., 2015). The much-publicised Course Signals project (Arnold & Pistilli, 2012) is an example of the kind of predictive work that was undertaken. Other examples include the 'early warning system' developed by Macfadyen and Dawson (2010), and the set of Australian case studies discussed by West (2015). A central focus of this work was on using the power of data and analytics to make predictions about student learning trajectories and possibly intervene with the intention of generating better outcomes (Clow, 2013). There has been some conjecture about how effective these projects have been. There are clear limitations of pure empirical, prediction-making approaches (e.g. Gašević et al., 2015). However, there is evidence to suggest that these kinds of predictions have proven useful in helping institutions help students, particularly in university settings, despite their fairly minimal use of educational theory (Ferguson et al., 2016). As such, these efforts could be described as deliberately pragmatic in nature.

While initiatives such as Course Signals (Arnold & Pistilli, 2012) no doubt paved the way for learning analytics to become what it is today, there emerged key questions that necessitated taking stock of where the field was headed. This early predictive work relied heavily on the use of demographic and behavioural data. Some (e.g. Lodge & Lewis, 2012), described this as a neo-behaviourist (see Skinner, 1954 for an overview of behaviourism related to education) way of thinking about what learning is and how it works. While innovative, there was a lot of emphasis in these early projects on the behaviour of students without as much thought about what was going on in their minds, what emotions they were experiencing and what their subjective experience of the learning environment were. Furthermore, as was described by Knight and Buckingham-Shum (2017), attempting to make any such predictions necessitates decisions about what is being measured, how it is being measured, what the purpose of the measurement is, and who the measurement is ultimately for. The kinds of data that were available and the lack of integration between theory and practice allowed for distal inferences only to be made. These inferences can be misleading or incorrect without an adequate means of interpretation (Gasevic et al., 2016). Of course, this is true of psychological measurement more broadly; it is not possible to observe and measure what occurs in students' minds directly. Learning analytics applications have historically tended to rely upon correlational data to make inferences about phenomena that cannot be observed directly (see also Lodge et al., 2017). This meant that the field needed to move beyond the early focus on what could be done to more carefully consider what predictions mean for real humans in real learning contexts.

In a nutshell, the criticisms of the initial approaches to the use of data science in education laid a foundation for the field to move beyond the behaviourist-type, strongly empirical basis, that may or may not have been adopted by default. There are numerous reasons why a purely empirical approach to mining and analysing data about learning in the absence of a systematic way of making sense of those data is problematic (see Wise & Shaffer, 2015). In order to mature, research and application of learning analytics needed to integrate more theory into the projects that were being undertaken (Wise & Shaffer, 2015). A corresponding shift, reflecting a recognised challenge in educational technology implementation more generally, has been to focus on the implementation, integration, and impact of tools in context rather than out of the lab. This shift is understandable and desirable given that the ultimate aim is to support a positive impact on learning.

Understandably, another area that has provided connections across many researchers in learning analytics since its inception is the issue of ethics (see e.g. Prinsloo & Slade, 2013). Although there are several theoretical traditions (such as critical and social theories: see Murphy, 2022 for an overview) that underpin the concerns about the ethical use of data in education, all seem to share common concerns about transparency, justice, equity and generally not being unscrupulous. The focus of this research has ranged from institutional policy (Prinsloo & Slade, 2013), to design processes for ethical learning analytics (Ahn et al., 2021), questions about dilemmas in practice (Kitto & Knight, 2019), to questions of justice (Wise et al., 2021). In places, this focus has led to the adoption of an instrumental approach (how can we implement standards for privacy?), while other work has addressed the fundamental values of learning analytics, and its position in wider systems of education (and society).

Education and the understanding and enhancement of student learning are not like the hard sciences, there are areas of conceptual murkiness, contested values, fundamental disagreements about appropriate metaphysical and epistemological assumptions, and an inability to directly observe and measure most of the key phenomena of interest (see Biglan, 1973 for a taxonomy of the characteristics of different fields). While data and analytics undoubtedly provide great potential for shedding light on learning processes and creating a foundation for enhancing student learning, these challenges mean that substantial theorisation is required to bridge the gap between the insights provided by data and what is occurring in the minds of real students in real educational environments. The quantitatively heavy approaches that were used in early learning analytics initiatives demonstrated the importance of theory in understanding and intervening with student learning. This issue undoubtedly contributed to the evolution of learning analytics as an interdisciplinary endeavour, which has ultimately become a strength of the field. We will discuss this aspect further on as we delve into multidisciplinary in learning analytics.

1.3 Theoretical critique from *without* learning analytics

A further major theoretical trend that is worth highlighting in this brief survey, are a collection of critical perspectives, problematising the methods, applications, and implications of learning analytics within the context of wider concerns regarding datafication, structural inequalities, managerialism, and more in society. This stream of work tends to address the question of what is 'learning analytics' by looking not at the research area, but at the social and political situatedness of learning analytics and its potential for misuse and generating poorer outcomes for subgroups in a society that already struggles with bias and discrimination. This commentary is 'without' learning analytics both in the sense that a line of critique has come from those outside the field, but also in the important sense that it reflects the field not only as a community of scholars, but also part of a wider system of technologies in society, including commercial 'learning analytics' tools. This work is an important reminder that

> "*the learning problems we identify and choose to work on are never blank slates; they embed societal structures, reflect the influence of past technologies; and have previous enablers, barriers and social mediation acting on them. In that context, we must ask the hard questions: What parts of existing systems is our work challenging? What parts is it reinforcing? Do these effects, intentional or not, align with our values and beliefs? In the end what makes learning analytics matter is our ability to contribute to progress on both immediate and long-standing challenges in learning, not only improving current systems, but also considering alternatives for what is and what could be.*" (Wise, Knight & Ochoa, 2021 p.1)

A feature of this commentary is that it serves to illustrate that there is not a singular, overarching view held within learning analytics. Indeed, even individual researchers acknowledge tensions and limitations in the field and their work. To give an example, Selwyn's (2019) paper on "What's the problem with learning analytics?" – an invited piece in *the Journal of Learning Analytics* based on his Learning Analytics and Knowledge Conference keynote in 2018 – fits the mould perfectly, situating Selwyn as the 'outsider', with four respondents (see Buckingham Shum, 2019). As Prinsloo's (2019) commentary on the piece in the same issue notes, this rhetorical move is intended to provoke and question. Clear among the four published responses in the journal issue is the receptiveness to this critique, and recognition of many of its themes in the field particularly regarding ethics (Ferguson, 2019; Prinsloo, 2019), although noting “many in the community resonate more with the same concerns, values, and vision forward that Selwyn has offered than they do with the characterization of the field that seems implicit in the writing” (Rose, 2019, pg. 31). As the authors of the commentary articles note, work in the field recognises the challenges of reduction as a means to abstraction and modelling while noting that qualitative approaches can be equally susceptible to error, and that models may help us *avoid* errors in interpretation: “Statistical aggregation is paradoxical. It defies common sense: by discarding information, we also gain information. Statistics is like that.” (Essa, 2019, pg. 37). All respondent authors also noted the diverse approaches in the field, highlighting the role of ‘multivocality’ in the field’s history and ongoing conversation “we should continue among ourselves to value our great diversity and work towards better understanding and appreciation of the many differences in perspective, methodology, and theoretical backgrounds in our midst” (Rose, 2019, pg. 34). A challenge, then, is how to expand this dialogue beyond the field to include critics not only in the ‘outsider’ role, but as active participants in the field’s discourse.

Additional critiques of learning analytics are provided by other researchers. Beyond the problems of privacy and ethics that were discussed above, some of the most relevant critical themes covered include: the manner in which vendors shape the direction of the field and the affordances offered to educators (Gulson et al., 2022; Williamson, 2017); the manner in which the collection of data affects educational policies (Williamson, 2019; Gulson and Sellar 2019); critique of the notion of objective data that is free of bias (Carter and Egliston, 2021), and the way data moves through education systems (Howard et al., 2022).

It is important to realise that many of these problems are also addressed *within* the field of learning analytics, as the section above has demonstrated, which has adopted a critical stance towards educational data science since its inception. We believe that much of the disconnect between the stance of critical data studies (CDS; see Iliadis & Russo, 2016 for an overview) and that of learning analytics stems from a split in how the field of learning analytics is itself defined. As the commercial sector takes an increasing interest in the field, we increasingly see poorly thought out and impoverished models of learning being sold across the sector for their learning analytics solution. CDS is right to challenge these approaches, but it would be naive to assume that this *is* learning analytics. We challenge those critiquing the field to ensure that they are aware of the best practices being used (along with the worst) rather than just working to attack straw men. One of the best reasons to advance the use of theory in learning analytics is for its potential role in improving the data analysis that all too often arises from big data analysed from an atheoretical perspective (Wise and Schaffer, 2015).

1.4 Theory used in learning analytics

Early work held great value in setting the foundation for the large and growing body of modelling and interventions that have come to make up the field of learning analytics. As the field has evolved, so have the connections with, and divergences from, other, similar fields.

Increasingly, learning analytics has become more sophisticated in the theoretical conceptualisation that it has adopted from various disciplinary contexts, and this has often resulted in more robust computational approaches. A central part of this transition has centred upon the way in which theories - particularly from the learning sciences - have been built on, and contributed to, learning analytics work. An early example of this drive is provided by papers in the special section of the Journal of Learning Analytics connecting theory to empirical work that was edited by Wise and Schaffer (2015). Importantly in their editorial Wise and Schaffer note that the way in which a theory is operationalised by a learning analytics effort matters, and that very little work has been completed on this important topic to date. Where are the general methods for turning a theory into an analytics method? This question is made more challenging when there is no general consensus about the appropriate theoretical lens for understanding the educational constructs being analysed.

In the most extreme examples, this can lead some researchers to claim that they are investigating problems and issues that nobody has looked at before, when, in fact, there is a body of research on the issue spanning decades and sometimes over a century. For example, and without wanting to pick on any discipline or individuals, a team of computer scientists might claim that people have not looked at a quantitative aspect of memory or attention in learning before while being completely unaware that the issue has been a topic of extensive research in experimental psychology since the 1960s. The reverse can also occur with psychological scientists being unaware of the algorithmic research of critical importance to their work in computer science. As has been described elsewhere, particularly in other multidisciplinary domains, such as the learning sciences (e.g. Palghat et al., 2017), bringing people from different disciplinary traditions together to work on a common problem is a complex and difficult endeavour. Such is the case with learning analytics. There is no commonly agreed-upon theoretical foundation. Individuals and teams will tend to draw on the theory that is most familiar to them, either through their own disciplinary upbringing or via a more interdisciplinary understanding that has evolved over their lives and careers. Some of the assumptions that are inherent within these traditions are mutually exclusive and, therefore, incompatible. At the extreme ends of the mismatches, a cognitive neuroscientist will likely take a reductionist view of learning that is biologically focused. At the other end of the spectrum, a sociologist will reject any reductionist or positivist notion and instead take a more relativist and situated view of learning (Palghat et al., 2017).

Despite these observations, there are many examples of projects and innovations in the area that are deeply situated within clear philosophical and methodological traditions. One example is the research employing social network analysis (SNA) theory (e.g. Dawson et al., 2010). SNA moved beyond purely data-driven approaches to systematically infer meaning from the connections between learners. This work has matured over the years and is now explicitly discussing the links to theory (Chen & Poquet, 2022).

Links have also been attempted between learning analytics and theories and ideas that fit within the broad areas of educational psychology and the learning sciences (see Rosé, 2018). These linkages span a wide range of activities from studies that utilise the extensive body of research on self-regulated learning (Lodge et al., 2019) to what has come to be known as multimodal learning analytics (Ochoa, 2017). The latter has moved beyond traditional educational psychology by incorporating methodologies that are more akin to psychophysiology (Cacioppo et al., 2007). Along with the closer linkages between these areas come established theories about learning, many of which can now be tested empirically by learning analytics-based methods. Thus, with the rise of learning analytics it is now becoming possible to use data collected “in the wild” to test competing theories, and perhaps to resolve long-running debates. For example, with the emerging interest displayed by many learning analytics researchers in self-regulated learning it may become possible to empirically determine which of the many different theories on this topic (Panadero, 2017) are the most robust, and most applicable to particular situations.

While there has been much emphasis on the use of quantitative demographic and behavioural data in learning analytics, it is not true to say that all research has been in this vein. Data commonly used in the field now includes a range of other sources and, increasingly, qualitative data. These new data sources raise issues of sensemaking and interpretation that sound theoretical models would help to resolve. Perhaps some of the earliest work in this regard used discourse analysis to make sense of comments and other forms of writing produced by students. For example, De Liddio and colleagues (2011) described an approach for using sociocultural discourse analysis to analyse annotation and deliberation text generated by learners. Analysing text-based data necessitated the use of complex theory, such as that of sociocultural discourse, to provide a path for making sense of these data. As is the case with quantitative, behavioural data, there is also an inferential gap between what is written and what is going on in the minds of students (Gibson, Kitto & Bruza, 2016). In this instance, linguistics provides some guidance about how to make sense of these data and what actions can be taken on the basis of them. Discourse analysis applied in this manner is an example of theoretical foundations being ‘borrowed’ or adapted from other research fields. This kind of adapting of theory has been a common feature of much of the research in learning analytics to date.

Emerging systematic reviews of theory in learning analytics also provide some insights about how the field is developing. For example, Wong et al. (2019) claim to review how learning theories and learning analytics could be integrated in educational research, finding that a clear majority of papers drew on the topic of self-regulated learning. However, their search string of “learning analytics” AND (“stud* success” OR “achievement”) means that they only explored a very specific subset of learning analytics and their results may not generalise. However, their results were recently supported by Wang et al. (2022), who attempted to trace the rise of theory since 2016 in the use of clickstream data, utilising "log analysis" and "education" to define a subset of learning analytics and then classifying the 37 papers included in their study for the type of educational construct returned. Again, self-regulated learning was overwhelmingly the dominant theory applied to the data. We note however that both of these reviews reflect the contested theorisation of learning analytics itself, in their selection of search terms that illustrate a particular perspective on, and therefore provide a particular lens to, learning analytics that may omit various important papers. In particular, the emerging studies linking qualitative data to educational constructs were not captured by these search terms. As such, they cannot be considered definitive in their tracing of the rise of theory across the field, and a more complete study is required to draw any strong conclusions on the theories drawn on in learning analytics to date.

Right at the dawn of the field, it was recognised that there was a very strong risk for learning analytics to encourage a return to behaviourism in education (Long & Siemens, 2011: Lodge & Lewis, 2012). However, it would seem more accurate to describe research in learning analytics as underpinned by constructivist notions, or, at least, post-positivist, given its evolution since then. As the years passed, it has become less common to see purely data-driven approaches in the field. Indeed, it could be argued that the division between learning analytics and educational data mining revolves around a focus on technical discovery (educational data mining) or processes of individual learning (learning analytics) (Baek and Doleck, 2021; Lemay, Baek, Doleck, 2021; Liñán, Pérez, 2015; Dormezil, Khoshgoftaar & Robinson-Bryant, 2019). There seems to be broad recognition emerging of the complexities associated with modelling learning and education. In many cases, this complexity is recognised explicitly. For example, Tsai et al. (2019), argue that complex systems theory is the most appropriate theoretical foundation for applied educational research and that the reasons behind non-adoption of learning analytics in higher education stem from a lack of agile leadership that is able to embrace this complexity.

In part due to the complexity of learning, increasingly the relationship between these fields or disciplines working in learning analytics has become tighter. Thus, theories with a grounding in the learning sciences, such as self-regulated learning, have informed the design, implementation and interpretation of educational data and analytics systems (Marzouk et al., 2016). As techniques like structural equation modelling and causal inference become more prominent in the field, we expect to see more studies that impose theoretical models upon datasets, an approach that can only lead to more explainable and actionable models. See as a good example of this phenomenon the recent special issue on linking learning analytics and assessment (Gašević et al., 2022) which provides a number of examples of this phenomenon. However, less work has been completed in the opposite direction (Lodge & Corrin, 2017), and we consider this a singular opportunity for the field: how can learning analytics start to inform, refine, and perhaps even choose between, educational theories?

1.5 Theory Arising from Learning Analytics

As the field has matured, theories have emerged directly from learning analytics work, characterised by their use of data science approaches in the service of theories of action (e.g. Argyris, 1993) for both researchers and practitioners in supporting learning. Insofar as there is a unified view of the field, that view - certainly as expressed by the conference and journal calls - is of a field that respects the “human in the loop” in applying analytics to learning (Lang et al., 2022). As those authors note, this provides a guiding line for our work, to investigate how data and analytics can augment human capacities, in which research contributions may come from gaining a deeper understanding of those interactions (Lang et al., 2022). These gains in knowledge - contributions to theory - may be in a deeper understanding of learning processes through data, new analytic approaches for probing data, novel understanding of data-informed feedback that has an impact on learning, and indeed design processes to inform all of this. In taking this approach, the field aims to situate its work as ‘problem-centric’ not ‘tool-centric’ (Wise et al., 2021), with a clear eye to practice aligned research that takes a design approach (Wise et al., 2020).

As an example, the gap between the data and the reality of learning in educational contexts brought about an approach to theorisation from learning analytics. Numerous researchers (e.g. Lockyer et al., 2013; Mor et al., 2015), made the argument that learning analytics needs to be integrated with learning design (see Maina et. al., 2015). Key to this argument is that decontextualised data is only ever going to have limited potential for action without a clear understanding of the goals or intent of the lesson or program of study (Lockyer et al., 2013). Adding learning design to the equation opened up opportunities for more deeply understanding the processes that students go through, providing a richness and depth to the kinds of inferences that could be made. For example, Bakharia and colleagues (2016) offer a means of connecting learning analytics with learning design through an understanding of the pedagogical intent of learning activities. Macfadyen et al. (2020) argue that learning design provides an 'interpretive pedagogical framework' that can act as a bridge to connect through what Knight et al. (2014) previously described as the 'middle space' between learning and analytics. In this way, learning design acts as a conduit for learning analytics to the long and deep history of theory in education.

A second key area in which learning analytics has spawned new understanding is in the combining of different forms of data. The increased synthesis of quantitative and qualitative data in learning analytics has arguably been taken a step further through what has come to be known as quantitative ethnography (see Shaffer, 2017). Shaffer describes quantitative ethnography as a means of treating big data sets as a form of cultural discourse. By blending quantitative and qualitative data, he argues that a deeper or, as he refers to it, 'thick' description of patterns in data can be achieved. With an annual Quantitative Ethnography Conference now taking place, this synthesis of big data and grounded, qualitative approaches is well established beyond the realm of education. Quantitative ethnography brings with it powerful theoretical tools for enhancing the understanding and enhancement of learning in real, complex educational environments. Quantitative ethnography represents both a way of making meaning of quantitative data by drawing on qualitative components and one of the key theoretical trends in learning analytics over time.

1.6 Concluding Remarks: The 'hard problem' of learning analytics?

As is hopefully evident in this chapter, the field of learning analytics has evolved from small-scale projects examining what could be done through to deep ethical discussions about whether it should be done, and an ongoing drive towards the integration of theory and data. Spanning these questions are projects, initiatives, and debates that view the use of data and analytics in education through many disciplinary, metaphysical, epistemological and ontological perspectives. It is perhaps not surprising that there is deep and persistent uncertainty in the community about core aspects associated with what learning is, what various forms of data mean, what kinds of inferences can be made and what is ethical and moral action on the basis of the data we now have. The field of learning analytics is inherently interdisciplinary and has often been described as a bricolage field (Siemens, 2014) that requires people to work in a `middle space' between the learning and analytical sciences (Knight et al., 2014). However, we have shown here that as the field matures through more rigorous attention to theory, it is gradually moving towards a more transdisciplinary and coherent perspective.

Along with these intersecting strands, individuals, lab groups and teams of collaborators all bring to the field their own disciplinary and scholarly traditions. These traditions, in turn, have provided the foundation of the theories that underpin the research occurring in learning analytics. As an applied field working in the complex context of education, having a variety of theories and world views to apply to the issues that students and teachers experience is broadly an advantage (despite the problems we have described here). We characterise these intersecting strands of activity into four groupings, theorising of learning analytics, theoretical critiques of learning analytics, theory use in learning analytics, and theories arising from learning analytics. We will provide a brief overview of each:

Theorising of learning analytics: This strand of activity has a few threads that have ebbed and flowed over time. The core of these threads is theorising about what learning analytics is, how it relates to other fields and disciplines and what kinds of impacts it has and is aiming for. Another way to conceptualise work in this area is theories *of* learning analytics, aiming to delineate learning analytics from related fields.

Theoretical critique about learning analytics: This strand of activity has come both from within and without the field, with a range of contributions to theory around systemic features of learning analytics, including ethics, implementation, methodological challenges, and broader socio-political challenges. Much of this work applies both to learning analytics, and broader fields of research and practice.

Theory use in learning analytics: As we have described in this chapter, research in learning analytics has drawn on a set of theoretical ideas, often arising from the learning sciences and related disciplines, as well as work in computer and cognitive sciences. Understanding this uptake of theory, and contributions back to established theory, is an important part of situating the field and its contributions.

Theory arising from learning analytics: The final intersection of theory and learning analytics is in the distinctive contributions to theory that the field makes. While the field is evolving, insofar as there is firm ground it seems to be in the conceptualisation of 'human in the loop' approaches to the intersection of data and learning, and the development of methodologies - informed by both social and computational sciences - to design tools to understand and support learning, and approaches to evaluate this impact. From this work, distinctive contributions and new theories of learning are emerging.

This chapter has attempted to outline some key developments in the history of learning analytics that have led the field to integrate theory into the mix more deeply. The trends that we have outlined here are among the most prominent. We do not claim to have provided an exhaustive list of the threads of work that have helped to create the theoretical foundations for learning analytics as they stand today. Nor could we do justice to the depth of the thinking and work completed in any of the areas we have mentioned. Instead, we have aimed to draw attention to the ways that the research and application of learning analytics has connected to theory. In doing so, we hope to provide a sense of the connections that have guided the field's maturing and suggest the possible paths forward for a deeper integration of theory and learning analytics, crucially, with an eye towards impact on learning, across learners.

1.7 References

Ahn, J., Campos, F., Nguyen, H., Hays, M., & Morrison, J. (2021). Co-Designing for Privacy, Transparency, and Trust in K-12 Learning Analytics. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 55–65. https://doi.org/10.1145/3448139.3448145

Anderson, C. (2008) The end of theory, will the data deluge makes the scientific method obsolete?, *Edge*, [Online] Retrieved from: http://www.edge.org/3rd_ culture/anderson08/anderson08_index.html

Argyris, C. (1993). *Knowledge for action*. Jossey-Bass.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270).

Baek, Clare & Doleck, Tenzin (2021) Educational Data Mining versus Learning Analytics: A Review of Publications From 2015 to 2019, *Interactive Learning Environments*, DOI: 10.1080/10494820.2021.1943689

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gasevic ́, D., Mulder, R. et al. (2016). A conceptual framework linking learning design with learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 329–338). ACM.

Balacheff, N., & Lund, K. (2013, April). Multidisciplinarity vsvs. Multivocality, the case of" Learning Analyytics". In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 5-13).

Biglan, A. (1973). The characteristics of subject matter. *Journal of Applied Psychology, 57*, 195–203

Buckingham Shum, S. (2019). Critical Data Studies, Abstraction & Learning Analytics: Editorial to Selwyn's LAK keynote and invited commentaries. *Journal of Learning Analytics*, *6*(3), 5-10.

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. (2007). *Handbook of Psychophysiology*. Cambridge, United Kingdom: Cambridge University Press.

Calvet Liñán, L., Juan Pérez, Á.A. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *International Journal of Educational Technology in Higher Education* 12, 98–112 (20//doi.org/10.7238/rusc.v12i3.2515

Carter, M., & Egliston, B. (2021). What are the risks of virtual reality data? Learning analytics, algorithmic bias and a fantasy of perfect data. *New Media & Society*, 14614448211012794.

Chen, B., & Poquet, O. (2022). Networks in Learning Analytics: Where Theory, Methodology, and Practice Intersect. *Journal of Learning Analytics*, *9*(1), 1-12.

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Shum, S. B., & Fisher, J. (2015). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Australia: Department of Education.

Dawson, S., Bakharia, A. & Heathcote, E. (2010). SNAPP: Realising the affordances of real-time SNA within networked learning environments. *Paper presented at Networked Learning–Seventh International Conference*, Aalborg, Denmark.

De Liddo, A., Shum, S. B., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2011, February). Discourse-centric learning analytics. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 23-33).

Dormezil, S., Khoshgoftaar, T. & Robinson-Bryant, F. (2019). Differentiating between educational data mining and learning analytics: a bibliometric approach, in *LABBEC Workshop (Learning Analytics: Building Bridges Between the Education and the Computing Communities)*, 1–6.

Elias, T. (January 2011). *Learning Analytics: Definitions, Processes and Potential* (PDF). Unpublished Paper: 19. S2CID 16906479

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5-6), 304-317.

Ferguson, R. (2019). Ethical Challenges for Learning Analytics. *Journal of Learning Analytics*, *6*(3), 25–30. https://doi.org/10.18608/jla.2019.63.5

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-71.

Gasevic, D., Dawson, S., Rogers, T. & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, Education, 28*(C), 68–84. http://doi.org/10.1016/j.iheduc.2015.10.002

Gašević, D., Greiff, S., & Shaffer, D. W. (2022). Towards Strengthening Links between Learning Analytics and Assessment: Challenges and Potentials of a Promising New Bond. *Computers in Human Behavior*, 107304.

Gibson, A., Kitto, K., & Bruza, P. (2016). Towards the discovery of learner metacognition from reflective writing. *Journal of Learning Analytics*, *3*(2), 22-36.

Gulson, K. N., & Sellar, S. (2019). Emerging data infrastructures and the new topologies of education policy. *Environment and Planning D: Society and Space*, *37*(2), 350-366.

Gulson, K. N., & Witzenberger, K. (2022). Repackaging authority: artificial intelligence, automated governance and education trade shows. *Journal of Education Policy*, *37*(1), 145-160.

Howard, S., Swist, T., Gasevic, D., Bartimote, K., Knight, S., Gulson, K., Apps, T., Peloche, J., Hutchinson, N., & Selwyn, N. (2022). Understanding educational data journeys: Where are we going, what are we taking and making? *Computers & Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2022.100073

Iliadis, A. & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society 3*(2): 1–7.

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, *50*(6), 2855-2870.

Knight, S., & Buckingham Shum, S. (2017). Theory and Learning Analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gaševic (Eds.), *The Handbook of Learning Analytal[yt]ics* (1st ed., pp. 17–22). Alberta, Canada: Society for Learning Analytics Research (SoLAR). Retrieved from http://solaresearch.org/hla-17/hla17- chapter1

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics, 1*(2), 23-47. https://doi.org/10.18608/jla.2014.12.3

Lang, C., Wise, A. F., Merceron, A., Gaševic, D. & Siemens, G. (2022) "What is Learning Anallyticss?." 10.18608/hla22.001

Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, *2*, 100016.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439-1459. https://doi.org/10.1177/0002764213479367

Lodge, J. M. & Corrin, L. (2017). What data and analytics can and do say about effective learning. *Nature: npj Science of Learning*, *2* (1), 4-5. http://dx.doi.org/10.1038/s41539-017-0006-5

Lodge, J. M. & Lewis, M. J. (2012). Pigeon pecks and mouse clicks: Putting the learning back into learning analytics. In M. Brown, M. Hartnett & T. Stewart (Eds.), Future challenges, sustainable futures. Proceedings ascilite Wellington 2012.

Lodge, J. M., Alhadad, S. S. J., Lewis, M. J. & Gašević, D. (2017). Inferring learning from big data: The importance of a transdisciplinary and multidimensional approach. *Technology, Knowledge & Learning, 22*(3), 385-400. https://doi.org/10.1007/s10758-017-9330-3

Lodge, J. M., Panadero, E., Broadbent, J. & de Barba, P. (2019). Supporting self-regulated learning with learning analytics. In J. M. Lodge, J. C., Horvath & L. Corrin (eds.).

Learning analytics in the classroom: Translating learning analytics research for teachers. Abingdon, UK: Routledge.
Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review* September/October, 31-40.
Macfadyen, L. P., Lockyer, L., & Rienties, B. (2020). Learning design and learning analytics: Snapshot 2020. *Journal of Learning Analytics, 7*(3), 6-12. https://doi.org/10.18608/jla.2020.73.2
Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and Education, 54*(2), 588-599.
Maina, M., Craft, B. & Mor, Y. (2015). *The art & science of learning design*. Springer.
Marzouk, Z., Rakovic, M., Liaqat, A., Vytasek, J., Samadi, D., Stewart-Alonso, J., Ram, I., Woloshen, S., Winne, P.H. & Nesbit, J. C. (2016). What if learning analytics were based on learning science?. *Australasian Journal of Educational Technology*, *32*(6).
Mor, Y., Ferguson, R., & Wasson, B. (2015). Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology, 46*(2), 221-229. https://doi.org/10.1111/bjet.12273
Murphy, M. (2022). *Social theory and education research (2nd ed.)*. Routledge.
Ochoa, X. (2017). Multimodal learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 129–141). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.011
Palghat, K., Horvath, J. C., & Lodge, J. M. (2017). The hard problem of 'educational neuroscience'. *Trends in Neuroscience & Education, 6* (1), 204-210. http://dx.doi.org/10.1016/j.tine.2017.02.001
Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 422.
Perrotta, C., Gulson, K. N., Williamson, B., & Witzenberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education*, *62*(1), 97-113.
Prinsloo, P. (2019). Commentary on Neil Selwyn's LAK18 Keynote Address. *Journal of Learning Analytics*, *6*(3), 20–24. https://doi.org/10.18608/jla.2019.63.4
Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 240–244. https://doi.org/10.1145/2460296.2460344
Rienties, B., Køhler Simonsen, H. & Herodotou, C. (2020) Defining the boundaries between Artificial Intelligence in Education, Computer-Supported Collaborative Learning, Educational Data Mining, and Learning Analytics: A need for coherence. *Frontiers in Education.* 5:128. doi: 10.3389/feduc.2020.00128
Rosé, C. P. (2018). Learning analytics in the learning sciences." In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann (eds.) *International Handbook of the Learning Sciences*, 511–519. Routledge.
Rose, C. P. (2019). Monolith, Multiplicity or Multivocality: What do we stand for and where do we go from here?. *Journal of Learning Analytics*, *6*(3), 31–34. https://doi.org/10.18608/jla.2019.63.6
Selwyn, N. (2019). What's the problem with learning analytics?. *Journal of Learning Analytics*, *6*(3), 11-19.
Selwyn, N. (2020). Re-imagining 'Learning Analytics'… a case for starting again?. *The Internet and Higher Education*, *46*, 100745.
Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.
Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400. https://doi.org/10.1177/0002764213498851
Siemens, G. (2014). The journal of learning analytics: Supporting and promoting learning analytics research. *Journal of Learning Analytics*, *1*(1), 3-5.

Siemens, G., & Baker, R. S. D. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254).

Skinner, B.F. (1954) The science of learning and the art of teaching. *Harvard Educational Review, 24,* 86–97.

Society for Learning Analytics Research (SoLAR; n.d.). *What is learning analytics?* Retrieved from solaresearch.org/about/what-is-learning-analytics/

Tsai, Y. S., Poquet, O., Gašević, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges and opportunities. *British Journal of Educational Technology*, *50*(6), 2839-2854.

Williamson, B. (2017). Learning in the ‘platform society’: Disassembling an educational data assemblage. *Research in Education*, *98*(1), 59-82.

Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. *British Journal of Educational Technology*, *50*(6), 2794-2809.

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior, 112*, 106457.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics, 2*(2), 5-13. https://doi.org/10.18608/jla.2015.22.2

Wise, A. F., Knight, S., & Ochoa, X. (2018). When are learning analytics ready and what are they ready for? *Journal of Learning Analytics*, *5*(3), 1-4.

Wise, A. F., Knight, S., & Ochoa, X. (2021). What Makes Learning Analytics Research Matter. *Journal of Learning Analytics*, *8*(3), 1-9. https://doi.org/10.18608/jla.2021.7647

Wise, A. F., Sarmiento, J. P., & Boothe Jr, M. (2021). Subversive learning analytics. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 639-645).

Wong, J., Baars, M., Koning, B. B. D., Zee, T. V. D., Davis, D., Khalil, M., Houben, G. & Paas, F. (2019). Educational theories and learning analytics: From data to knowledge. In *Utilizing learning analytics to support study success* (pp. 3-25). Springer, Cham.